Free Energies of Hydration from Thermodynamic Integration: Comparison of Molecular Mechanics Force Fields and Evaluation of Calculation Accuracy

VOLKHARD HELMS and REBECCA C. WADE*

European Molecular Biology Laboratory, Meyerhofstr. 1, 69117 Heidelberg, Germany

Received 13 December 1995; accepted 19 June 1996

ABSTRACT

Four commonly used molecular mechanics force fields, CHARMM22, OPLS, CVFF, and GROMOS87, are compared for their ability to reproduce experimental free energies of hydration (ΔG_{hydr}) from molecular dynamics (MD) simulations for a set of small nonpolar and polar organic molecules: propane, cyclopropane, dimethylether, and acetone. ΔG_{hydr} values were calculated by multiconfiguration thermodynamic integration for each of the different force fields with three different sets of partial atomic charges: full charges from an electrostatic potential fit (ESP), and ESP charges scaled by 0.8 and 0.6. All force fields, except for GROMOS87, give reasonable results for ΔG_{hydr} if partial atomic charges of appropriate magnitude are assigned. For GROMOS87, the agreement with experiment for hydrocarbons (propane and ethane) was improved considerably by modifying the repulsive part of the carbon–water oxygen Lennard–Jones potential. The small molecules studied are related to the chemical moieties constituting camphor ($C_{10}H_{16}O$). By invoking force-field transferability, we calculated the ΔG_{hydr} for camphor. With the modified GROMOS force field, a ΔG_{hydr} within 4 kJ/mol of the experimental value of -14.8 kJ/mol was obtained. Camphor is one of the largest molecules for which an absolute hydration free energy has been calculated by molecular simulation. The accuracy and reliability of the thermodynamic integration calculations were analyzed in detail and we found that, for ΔG_{hydr} calculations for the set of small molecules in aqueous solution, molecular dynamics simulations of 0.8–1.0 ns in length give an upper statistical error bound of 1.5 kJ/mol, whereas shorter simulations

*Author to whom all correspondence should be addressed.
E-mail: wade@embl-heidelberg.de

Journal of Computational Chemistry, Vol. 18, No. 4, 449–462 (1997)

CCC 0192-8651 / 97 / 040449-14

of 0.25 nm in length given an upper statistical error bound of 3.5 kJ/mol.

Introduction

Molecular hydration is of fundamental importance in chemical and biological processes. The solubility of a molecule is dependent on its free energy of hydration (ΔG_{hydr}). The binding affinity of one molecule to another is dependent on its ΔG_{hydr}. Thus, information about ΔG_{hydr} is required to predict the binding affinity of drugs to receptors, substrates to enzymes, guests to hosts, and so on. It is also crucial to understanding the hydrophobic effect. To make quantitative predictions, it is necessary to be able to calculate ΔG_{hydr} for organic molecules to high accuracy and this is the subject of this article. ΔG_{hydr} calculations also provide a fundamental calibratory test of molecular mechanics (MM) force fields and computational methods.

Different representations of the solvent have been used to study hydration processes.[1] ΔG_{hydr} values of organic solutes can be computed very accurately with continuum solvent models.[2,3] Continuum solvent models are computationally inexpensive compared to molecular simulation methods that treat solvent molecules explicitly. In general, good agreement between these two methods has been found for ΔG_{hydr} of organic solute molecules.[4–6] However, there are exceptions. For example, the ΔG_{hydr} of a single water molecule in aqueous solvent calculated by continuum methods and by molecular dynamics (MD) simulation was found to differ by ea. 8 kJ/mol as a result of nonlinear solvent effects that are not modeled in the continuum representation.[7] Continuum models can also give inaccurate ΔG_{hydr} in the case of solutes in which there is competition between intrasolute and solute–solvent hydrogen bond formation.[8] The present study is part of a larger project to calculate free energies associated with protein–ligand interactions. The protein simulation requires explicit consideration of protein dynamics and interactions with individual water molecules. It is desirable to use the same methods for all parts of a thermodynamic cycle so that systematic errors cancel and, therefore, we here employ MD simulations with explicit solvent molecules to compute ΔG_{hydr}.

There is now a considerable body of calculations of relative ΔG_{hydr} from molecular simulations with explicit solvent for organic molecules.[9–12] In these simulations, one solute molecule is perturbed into another and the relative free energy difference for the solvation of both molecules is calculated by thermodynamic perturbation (TP)[13] or thermodynamic integration (TI).* Perturbations can entail changes in atom types, bond lengths, and the number of atoms. On the other hand, the ΔG_{hydr} of a solute molecule with respect to a gaseous reference state can be calculated by introducing it into or removing it from the solvent while keeping its topology constant. Numerical problems while inserting or removing a particle, known as the ''origin singularity,'' may occur when the interaction of the perturbed particle becomes very small. These are more severe when slow-growth TI is used, and so these ΔG_{hydr} calculations have been done using TP,[14,15] or TI combined with TP.[16] These problems can be reduced by the techniques of bond shrinking,[17] EL decoupling, and using a nonlinear dependence on the perturbation parameter λ.[18] Recently, two similar techniques, termed separation-shifted potential scaling[19] and soft-core potential,[20] were developed that allow the origin singularity to be completely removed. This allows TI to be used alone for ΔG_{hydr} calculations. The robustness and versatility of TI, as well as the ability to estimate statistical errors, have been considerably improved by the multiconfiguration thermodynamic integration (MCTI) method.[21] We use this technique here in combination with separation-shifted potential scaling.

Camphor is the substrate of the enzyme cytochrome P450cam and binds to it in a buried protein cavity. We are interested in calculating binding free energies for this protein–ligand complex and this requires the calculation of ΔG_{hydr} for camphor to account for the desolvation of camphor on binding. Since the experimental ΔG_{hydr} of camphor is known, this calculation can also be used as a check of the parameterization of camphor. In test calculations with the original GROMOS87 parame-

*The free energy difference, $\Delta G_{0\rightarrow 1}$, between a system in states $\lambda = 0$ and $\lambda = 1$ can be calculated by performing N_λ simulations at intermediate λ values and computing the free energy by TP via $\Delta G_{0\rightarrow 1} = -k_B T \sum_{i=1}^{N_\lambda} \ln\langle \exp(-(H_{i+1} - H_i)/(k_B T))\rangle_i$, and by TI via $\Delta G_{0\rightarrow 1} = \sum_{i=1}^{N_\lambda} \langle \partial H(\lambda)/\partial\lambda\rangle_\lambda \Delta\lambda$.

ter set,[22] the computed ΔG_{hydr} of camphor favored hydration by 40 kJ/mol more than the experimental ΔG_{hydr}. It has been realized recently that GROMOS87 overestimates the solubility of alkanes in water[23] and this discrepancy has been resolved by making the Lennard–Jones (LJ) interaction between aliphatic carbons and water more repulsive. The same modified interaction parameter has been used in simulation studies of cyclodextrins[24] and proteins,[25] and an intermediate modification has been suggested recently,[26] but so far no systematic study of the influence of the interaction parameter has been published.

Camphor is a rather spherical molecule whose molecular volume (ea. 0.14 nm^3) corresponds to that of approximately five water molecules. It is about two thirds the size of biotin which, to our knowledge, is the largest molecule for which an absolute ΔG_{hydr} value has been calculated from MD simulations.[27] During the course of a perturbation calculation, the solvent has to rearrange considerably to fill the volume occupied by the solute. We will show that, with separation-shifted potential scaling, such transformations can be handled very smoothly. However, the use of relatively large molecules like camphor is not practical for fine tuning a force field for two reasons. First, solvent relaxation can be expected to be slower than for small molecules because a larger volume needs to be filled. This increases the equilibration time which cannot be used for analysis. Second, the simulation box necessary for camphor is larger than for small molecules which increases the computation time for each time step. Consequently, a few small model compounds were chosen and ΔG_{hydr} calculations were performed for these with the original GROMOS87 force field, with two variants of GROMOS87, and with three other well-known MM force fields: CHARMM22,[28] OPLS,[29] and CVFF.[30] Our aim was to identify similarities and differences between these force fields and to critically analyze the performance of GROMOS87 and the two GROMOS variants.

The four compounds, propane, cyclopropane, acetone, and dimethylether, are more or less rigid, and they represent different parts of the camphor molecule and related compounds. Choosing rigid solute structures eliminates the problem of proper sampling of the conformational degrees of freedom of the solute, and also avoids the problem that atomic point charges determined with the method of EL fit can vary considerably for different solute geometries.[31] Acetone was chosen as a model for the functional carbonyl group of camphor, and dimethylether was chosen because we are also studying the binding to cytochrome P450cam of camphor analogs that contain an ether moiety.[32] Propane and cyclopropane serve as models for the hydrophobic part of camphor, with cyclopropane representing the conformationally strained camphor body. Interestingly, the experimental ΔG_{hydr} of cyclopropane is more favorable than that of propane by 5 kJ/mol.

Of the currently available MM force fields, only the OPLS force field, and the recent revision of the AMBER force field[33] have been parameterized for the calculation of ΔG_{hydr} values. We chose OPLS as a representative member. In the force fields studied here, charge assignment is usually done with the charge template method. However, better defined charges are obtained by fitting the partial atomic charges to reproduce the molecular electrostatic (EL) potential calculated by quantum-mechanical methods (electrostatic potential fit (ESP)).[34–37] It was shown previously that a combination of OPLS nonbonded parameters and ESP charges can successfully reproduce the experimental ΔG_{hydr} of small organic molecules with diverse functional groups.[38] Therefore, we used ESP charges for the comparison of the different force fields, although we emphasize that ESP charges are not standard for any of the force fields studied. The relative weighting of EL and LJ forces varies in different MM force fields; i.e., the magnitude of the atomic charges varies. We did calculations with full, unscaled ESP charges, and with ESP charges sealed by 0.8 and 0.6, so as to cover the range of charge magnitudes typically encountered in these MM force fields. We can therefore make recommendations about the charge magnitude that give the best results for ΔG_{hydr} calculations with a specific force field type.

The calculation of the EL contribution to a solvation process is known to converge faster and more reliably than that due to hydrophobic interactions.[11,19] The creation or annihilation of a nonpolar cavity apparently involves larger rearrangements of the solvent surroundings than the creation or annihilation of a dipole or a net charge. Long simulation times are necessary to obtain reasonably accurate free energies for the dissolution of nonpolar solutes in aqueous solution.[39] These observations show that it is important to monitor the convergence and accuracy of free energy simulations carefully. It has been suggested that the inherent error in absolute ΔG_{hydr} values from typical 400-ps-long free energy perturbation calculations is on the order of 10%.[15] It would be highly

desirable if one could give an error estimate for a free energy perturbation calculation before starting the calculation or know how long a simulation needs to be in order to obtain results to a given accuracy. For the free energy calculations in this study, we will address the following questions: How great is the expected change in ΔG_{hydr} when simulation time is extended? How great is the deviation in ΔG_{hydr} calculated from simulations of the same system that were started from different starting configurations? How indicative is the statistical error calculated from the autocorrelation of $\partial H/\partial\lambda$? How indicative is the drift in $\partial H/\partial\lambda$ of the error in the total ΔG_{hydr}?

Details of the calculations are given in the Methods section. In the Results and Discussion section, we first present the results of the simulations with different force fields and discuss the tuning of the GROMOS carbon–water LJ parameter. Then, we address the accuracy of the simulations.

Methods

PARAMETRIZATION

Each compound was studied with seven force fields or force field variants: CHARMM22 in the commercial release; OPLS in combination with two solvent models (TIP3P and SPC/E); CVFF; GROMOS87; and two GROMOS variants (GROMOS–POL and GROMOS–690). In GROMOS–POL, the repulsive part of the LJ interaction (C12 parameter) between aliphatic carbons and water oxygens was changed from nonpolar (421 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$) to polar character (793 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$).[23] GROMOS–690 contains an intermediate modification of the same C12 parameter. The value of 690 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$ was suggested recently[26] and we also find this value to give the best fit with experimental data in calculations for ethane and propane (see below). Simulations were run independently for each force field.

All-atom models were used for all simulations except those with the GROMOS force field. Standard bonded parameters for each of the specific force fields were used for all compounds, except for camphor, for which the crystallographic coordinates of camphor in the complex with cytochrome p450cam were used (entry 2cpp in the Protein Data Bank[40]). The geometries of the model compounds and camphor were optimized at the HF/6-31G* level with the GAUSSIAN92 program [41] and partial atomic ESP charges were constructed with the CHELPG method.[36] These are given in Table I. The charges on aliphatic hydrogen atoms were averaged to reduce their conformational dependence. For the GROMOS force field, the charges of nonpolar hydrogens were summed into those of the nearest carbon united atoms. For most systems, three different sets of partial atomic charges were studied: unscaled ESP charges, and ESP charges scaled by 0.8 and 0.6. For camphor with the GROMOS87 and GROMOS–690 force fields, we also performed ΔG_{hydr} calculations with template charges of $+0.38/-0.38$ e assigned to the carbonyl group analogously to the carbonyl group of peptide bonds in the GROMOS87 force field. The charge on all the other atoms in camphor were set to zero, in accordance with the usual treatment of aliphatic moieties in the GROMOS87 force field.

TABLE I.
Computed ESP Partial Atomic Charges.

	Atom	Partial atomic charge [e]
Acetone	C2,C3	−0.374
	H	0.095
	C1	0.766
	O	−0.588
Dimethylether	C	0.095
	H	0.038
	O	−0.418
Propane	C1,C3	−0.211
	H1,H3	0.041
	C2	0.288
	H2	−0.056
Cyclopropane	C	−0.228
	H	0.114
Camphor	C1	0.12
	C2	0.56
	O	−0.57
	C3	−0.28
	H3exo / endo	0.08 / 0.09
	C4	0.04
	H4	−0.01
	C5	−0.14
	H5exo / endo	0.04 / 0.02
	C6	0.01
	H6exo,endo	0.0 / 0.0
	C7	0.42
	C8	−0.38
	H8	0.07
	C9	−0.43
	H9	0.09
	C10	−0.33
	H10	0.07

In these simulations, the carbonyl group was treated as one charge group, and the other atoms were defined as separate charge groups. In simulations with ESP charges, the molecules were defined as single charge groups.

An essential part of all MM force fields is the water model and we used a complementary water model for each force field: SPC/E water[42] for the three GROMOS variants; the flexible CVFF water model[43] for CVFF; and TIP3P water[44] for CHARMM. For OPLS, we used both TIP3P water and SPC/E water. The two versions of TIP3P water are different: the CHARMM version has LJ interactions assigned to the water oxygen and hydrogens; whereas, in the OPLS version, only the oxygen exerts LJ interactions.

SIMULATIONS

Versions 6.0 and 6.1 of the ARGOS program[45] were used for the MD equilibration and MCTI calculations.

A sharp group-based cut-off of 1.0 nm was used for simulations with all force fields. All interactions within a nonbonded cut-off of 0.8 nm were calculated every step, and long range interactions between 0.8 and 1.0 nm were updated every fifth step. The solvent–solvent nonbonded pairlist was updated every 10 steps and the solvent–solute list every 15 steps. All bondlengths were constrained with the SHAKE algorithm[46] with a relative tolerance on 10^{-8}. A time step of 2 fs was used. The temperature and pressure of the system were coupled to a temperature bath of 300 K and a pressure bath of 10^5 Pa with Berendsen thermostats[47] and coupling times of 0.4 ps and 0.5 ps, respectively. For the MCTI calculations, we used the separation-shifted potential scaling technique.[19] The potential scaling parameter, δ, was 0.05 nm^2, except for in the test calculations shown in Figure 1.

All compounds except camphor were solvated in a cubic periodic box of pre-equilibrated solvent with 2.4-nm box dimensions (456 to 457 solvent molecules), and the systems were energy minimized with 100 steps of steepest descent. After initial assignment of random velocities, 100 ps of MD equilibration at 300 K was performed for each system under NPT conditions, and interaction energies between the compound and solvent were calculated during the subsequent 100 ps. Camphor was solvated in a rectangular periodic box with dimensions of ca. 2.6 to 2.9 nm, so that there was a distance of at least 1.2 nm from any solute atom to the box exterior (577 solvent molecules for GROMOS, 852 for OPLS, 860 for CVFF, and 891 for CHARMM). The same equilibration protocol was used as for the smaller compounds.

The free energy perturbation studies were always started from an equilibrated system of one solute molecule in a box of solvent molecules. During a simulation, the nonbonded interactions between solute and solvent were switched off gradually. As the molecules have little conformational flexibility, intrasolute contributions were not included in the calculated ΔG_{hydr} values. The ΔG_{hydr} calculations were initially performed over $21 \times [1000 + 5000] \times 0.002$ ps = 252 ps; i.e., 21 simulations at equally spaced values of λ, each consisting of 1000 steps of equilibration followed by 5000 steps of data acquisition. This simulation stage will be referred to as the ''*initial simulation*.'' The statistical error in $\partial H/\partial\lambda$ in each window was calculated from the autocorrelation of the data.[48] Extension of the simulation time by a factor α leads to a decrease in the statistical error that is approximately $1/\sqrt{\alpha}$. In cases where the convergence plot for $\partial H/\partial\lambda$ versus data acquisition time indicated that the equilibration time was sufficient, the data acquisition part of the simulation was extended with the dynamic windowing technique.[49] This means that only those windows were extended in which either the statistical error in $\partial H/\partial\lambda$ exceeded 5 kJ/mol or the drift of $\partial H/\partial\lambda$ during the window exceeded 5 kJ/mol. Such windows were extended until both criteria were met or until a maximum of 30,000 steps. We use the notation $21 \times [1000 + (5000 < x < 30{,}000)]$ to represent these simulations because single windows can contain between 5000 and 30,000 data acquisition steps. For most simulations, equilibration and data collection times were extended to $21 \times [6000 + 15{,}000] \times 0.002$ ps = 882 ps and then with dynamic windowing to $21 \times [6000 + (15{,}000 < x < 30{,}000)]$ steps. The latter calculations will be referred to as ''*extended simulations*.''

The calculations for ethane and propane to optimize the C12 interaction parameter between the aliphatic carbons and the water oxygen consisted of 100-ps independent MD equilibration of each system, and ΔG_{hydr} calculations of $21 \times [6000 + (15{,}000 < x < 60{,}000)]$ steps.

Acetone, with the OPLS force field, the TIP3P water model, and unscaled ESP charges, was chosen as a test system to study the accuracy of the simulations because good agreement with experiment had been found previously.[50] To investigate

the sensitivity of the simulation to different starting configurations, we generated five starting positions for acetone by turning it through 90° or 180° around the *x*-, *y*-, and *z*-axes before solvating it in the box of water. Each system was equilibrated separately with 200 ps of MD, and ΔG_{hydr} was calculated with an *extended simulation*.

The same acetone test system was also used to study the effect of different values of the shifting parameter, δ, for the separation-shifted potential scaling that is applied to both LJ and EL interactions. By way of example, the scaled function for the LJ potential has the following form[19]:

$$V_{LJ} = (1 - \lambda)\left[C12/(r^2 + \delta\lambda)^6 - C6/(r^2 + \delta\lambda)^3\right]$$

where C12 and C6 are the repulsive and attractive LJ parameters, r is the distance between two interacting atoms, and λ is the coupling parameter. All MCTI calculations were started from the same equilibrated starting configuration and were *extended simulations*.

Results and Discussion

Figure 1A shows five ΔG_{hydr} calculations with different shifting parameter δ for the acetone test system. Without separation shifted potential scaling ($\delta = 0$), the free energy change for the process in which the solvent moves in to occupy the solute volume is mostly calculated in the last window and $\partial H/\partial\lambda$ converges very poorly in this window. For nonzero values of δ, this process takes place over a larger number of windows because the solute diameter is artificially decreased during the simulation. Therefore, the free energy curves can have very different shapes. For $\delta = 0.025$ nm, $\partial H/\partial\lambda$ is very steep in the last window and not well defined (-367 ± 15 kJ/mol after 60 ps of data collection in the last window). A similar behavior was previously observed for the annihilation of neon with $\delta = 0.01$ nm.[19] In that study, the optimal δ for neon was between 0.03 and 0.05 nm. In our case, the final ΔG_{hydr} results for the three simulations with $\delta = 0.05$ nm, $\delta = 0.075$ nm, and $\delta = 0.100$ nm are within 2.5 kJ/mol. Figure 1B displays the statistical error per window versus the coupling parameter λ. As δ is increased, the statistical error is reduced in the last windows and is distributed more evenly over all the windows. However, a value of $\delta = 0.10$ nm leads to a much larger statistical error in the free energy calculated

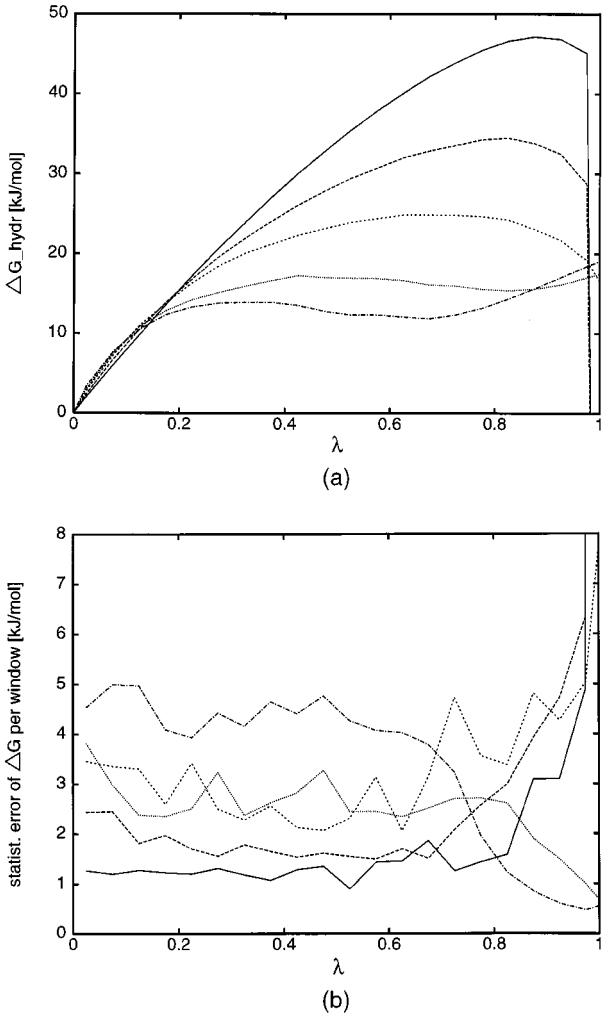

FIGURE 1. The effect of the potential shifting parameter, δ, is tested in ΔG_{hydr} calculations for the test system (acetone with the OPLS force field, TIP3P water model, and unscaled ESP charges). (A) shows the accumulated ΔG_{hydr} versus the coupling parameter, λ, and (B) shows the statistical error in each window. The lines correspond to $\delta = 0$ (——), 0.025 nm (– – –), 0.050 nm (- - - -), 0.075 nm (· · · ·), and 0.100 nm (- · - · -). The simulations are *extended simulations*. $\partial H/\partial\lambda$ does not converge to reasonable values in the last window of the simulations with $\delta = 0$ and $\delta = 0.025$ nm due to "origin singularities."

in the first few windows, which is undesirable. For this study, we used $\delta = 0.05$ nm because it results in a more conventional shape of the curve of free energy change versus λ and because we have previously found that this value worked well for perturbations of a water molecule.[51]

The C12 value for the carbon–water oxygen (OW) interaction for GROMOS–690 was deter-

mined from a number of calculations for propane and ethane where the C12 parameter was varied between 650 and 750 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$. The results are shown in Figure 2A for ethane and Figure 2B for propane. The calculated ΔG_{hydr} values seem to vary linearly with the C12 parameter. From linear regression, 702 and 672 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$ correspond to the experimental values[52] of 8.2 kJ/mol for propane and 7.7 kJ/mol for ethane. A value of 690 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$ is therefore a good compromise. The results for propane with GROMOS87 (−11.7 kJ/mol at 421 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$), and with GROMOS–POL (15.1 kJ/mol at 793 [kcal $Å^{-12}$ $mol^{-1}]^{1/2}$), lie on almost the same line (not shown). Such linear relationships could be usefully exploited in force-field parameterizations.[53]

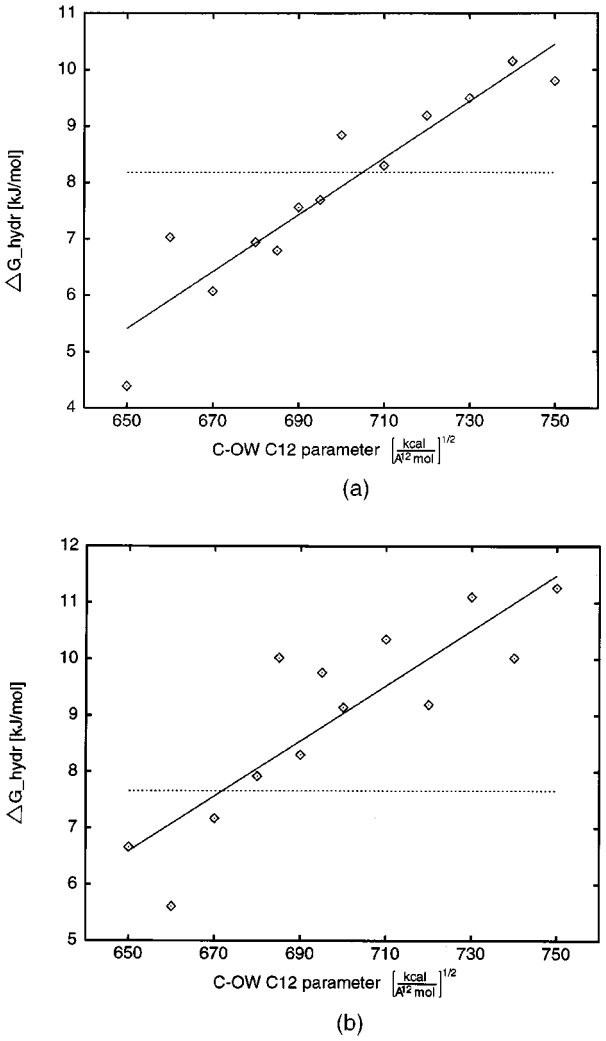

FIGURE 2. Calibration of the repulsive GROMOS C12 parameter for the LJ interaction between aliphatic carbon atoms and the SPC / E water oxygen. *Extended* MCTI calculations were performed for propane (A) and ethane (B). The horizontal dotted lines indicate the experimental values. The straight lines are linear fits to the calculated ΔG_{hydr} values.

COMPARISON OF DIFFERENT MOLECULAR MECHANICS FORCE FIELDS

Figure 3 A–H show the interaction energies between solute and solvent from MD simulations of acetone, dimethylether, propane, and cyclopropane in periodic boxes of water molecules. Figure 4 contains ΔG_{hydr} values from MCTI calculations that were started with configurations from these MD simulations. The LJ energy between solute and solvent is displayed for each compound. For acetone and dimethylether, we also show the EL energy and the total interaction energy as the sum of EL and LJ energies. The EL energies for propane and cyclopropane (not shown) are zero for GROMOS and small (< 0.3 kJ/mol for propane and -2.2 to -3.2 kJ/mol for cyclopropane) for the other force fields. No error bars are shown in the figures for clarity. The standard errors of the interaction energies are generally less than 0.5 kJ/mol for all simulations and the errors in the ΔG_{hydr} calculations are addressed in detail in the second part of this study.

The differences between the various force fields are more pronounced for the LJ energies than for the EL energies. The differences between the EL energies increase with the magnitude of the charges assigned (ca. 4 kJ/mol for 0.6 × ESP, ca. 10 kJ/mol for 0.8 × ESP, and ca. 15 kJ/mol for unscaled ESP charges). The LJ energies vary by as much as 23 kJ/mol, and the variability does not depend on the magnitude of the charges. The relative order of the force fields is similar for all four compounds. The EL energies for CVFF are more negative than for GROMOS87, and the LJ energies are more positive. Therefore, the total energies of CVFF and GROMOS87 are similar for acetone and dimethylether and the total CVFF energy is more positive for propane and cyclopropane. The CHARMM energies are in the same range as the CVFF and GROMOS87 energies, but they can be either larger or smaller in individual cases. The OPLS LJ energies are more positive than those with the three other force fields by ca. 10–20 kJ/mol. The OPLS interaction energies with SPC/E water are slightly more negative than with TIP3P water by ca. 1–2 kJ/mol. A noticeable difference only exists for acetone with 0.8 × ESP and with unscaled ESP charges where

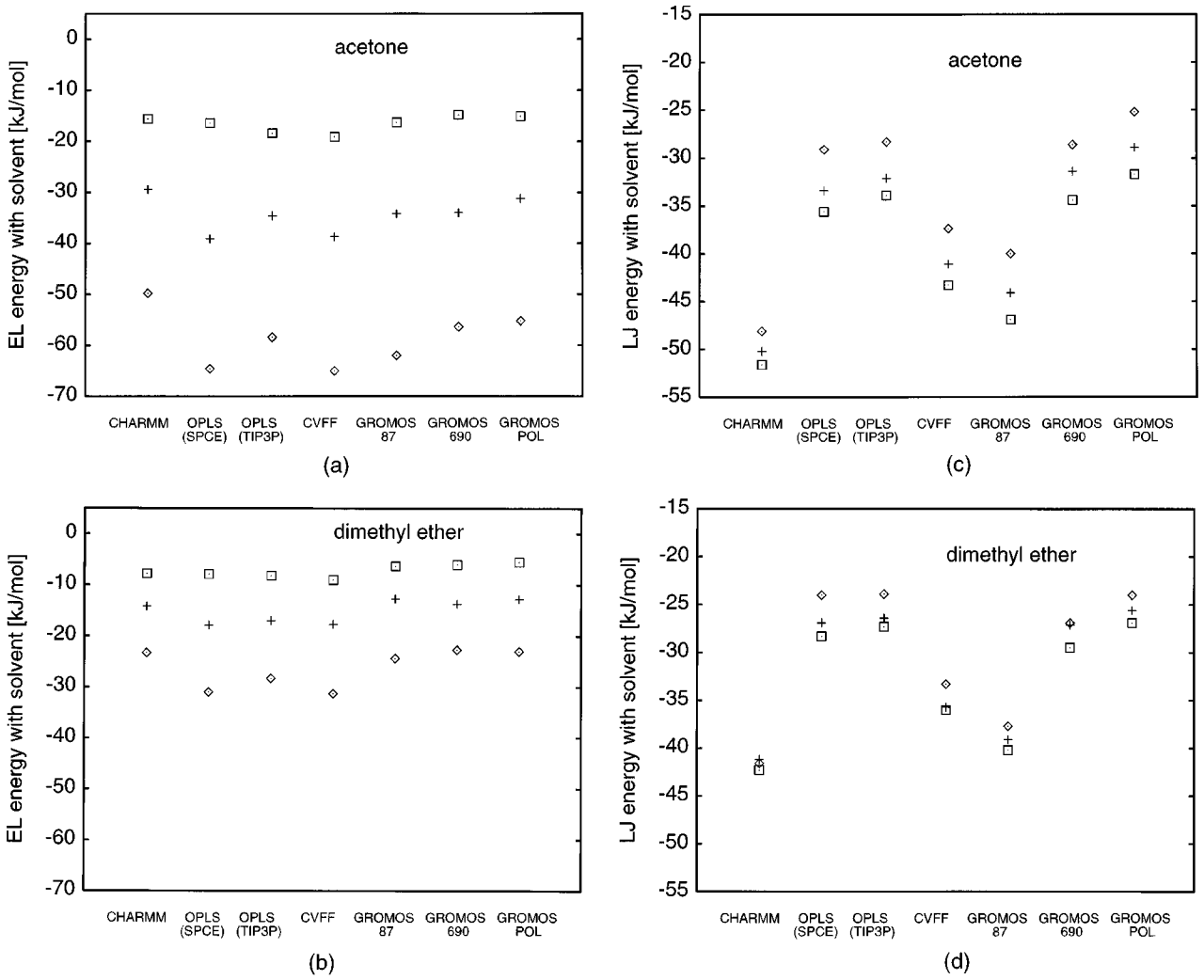

FIGURE 3. Solute–solvent interaction energies for acetone, dimethylether, propane, and cyclopropane during MD simulations with different force fields. The three columns show the EL contribution (a, b), the LJ contribution (c–f), and the total energy as the sum of the former two (g, h). Only the LJ energies are given for propane and cyclopropane. In general, each force field has three entries that result from independent MD simulation calculations for three sets of atomic charges: unscaled ESP charges (diamond); ESP charges scaled by 0.8 (+); and ESP charges scaled by 0.6 (square). Propane and cyclopropane were assigned zero charges with the GROMOS force fields, and a different symbol is used (×). Cyclopropane with the all-atom force fields and propane with OPLS–TIP3P were only simulated with unscaled ESP charges.

SPC/E is more negative by 5.8 and by 7.0 kJ/mol. The interaction energies for the two GROMOS variants are in the same range as the OPLS energies. GROMOS–POL naturally gives more positive interaction energies than GROMOS–690.

As can be expected from the small EL energies, the charge magnitude is not very important for aliphatic molecules, as long as the principle of neutral charge groups is maintained. The ΔG_{hydr} results for propane with unscaled and scaled ESP charges are reasonably similar for the all-atom force fields. All force fields give similar ΔG_{hydr} values for propane and cyclopropane which is inconsistent with the experimental difference of 5.3 kJ/mol favoring the solvation of cyclopropane. While both molecules are very similar in size and polarity, propane is considerably more flexible in vacuo than in solution, whereas the flexibility of cyclopropane does not change much. Propane can therefore be expected to be less soluble than cyclopropane due to the loss of intramolecular degrees of freedom. Intramolecular contributions to ΔG_{hydr}

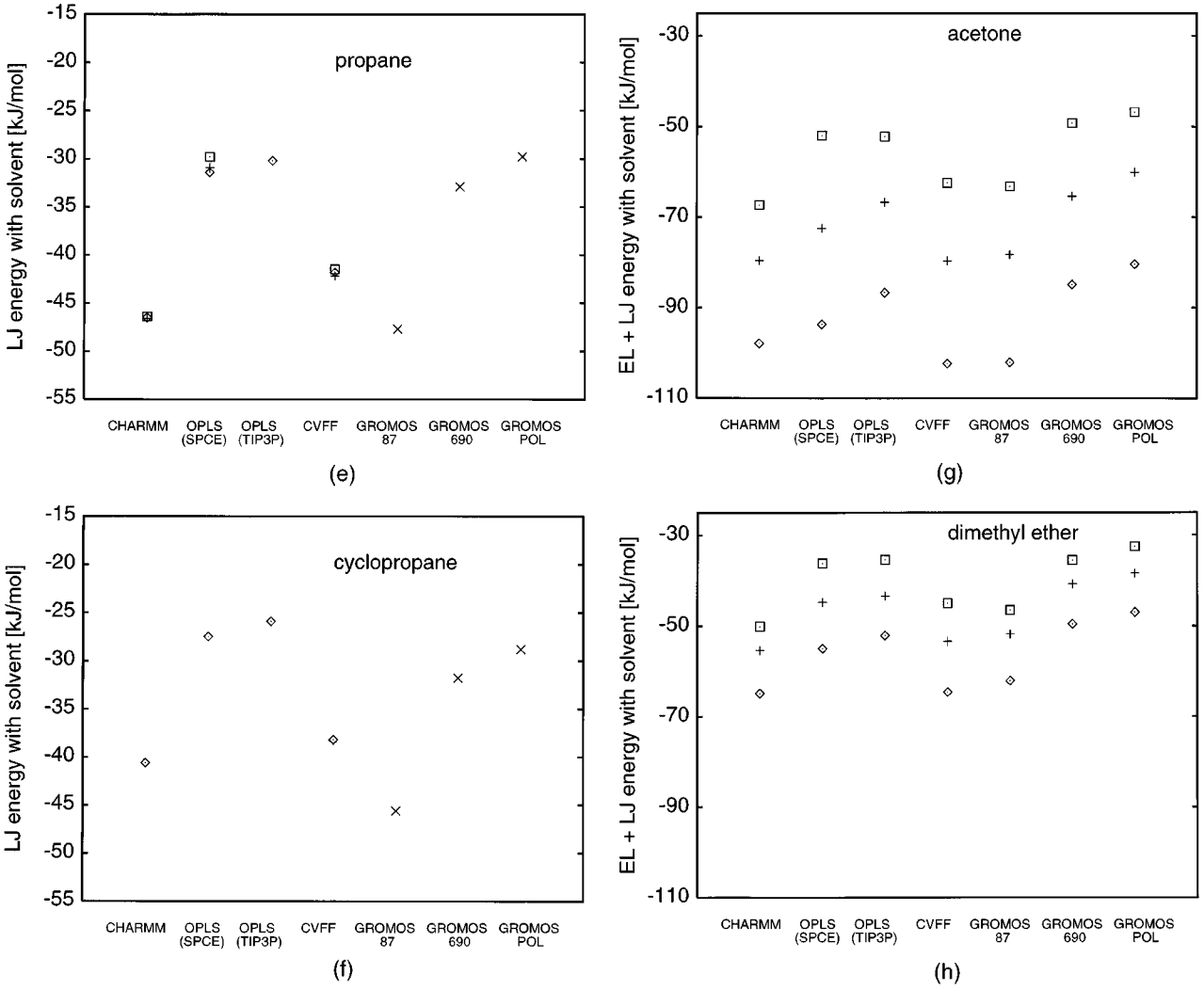

FIGURE 3. Continued.

were not considered in our simulations. As a general trend for the nonpolar compounds, the GROMOS87 force field is far too attractive by ca. 15–20 kJ/mol. The CHARMM and CVFF force fields give results closer to the experimental value, but they are also too attractive by ca. 5–10 kJ/mol. On the other hand, the OPLS results are slightly too unfavorable by 2–6 kJ/mol. The two GROMOS variants are more positive than GROMOS87 by ca. 15–20 kJ/mol (GROMOS–690) and ca. 20–25 kJ/mol (GROMOS–POL), and both variants give ΔG_{hydr} values that are in the range of the OPLS parameters.

The same trends are observed for the more polar compounds as for propane and cyclopropane. For these, however, agreement with experimental values is reasonable for all force fields if the magnitude of the atomic charges is considered as an adjustable parameter. The ΔG_{hydr} values for acetone and dimethylether for CHARMM and CVFF are within 3 kJ/mol of the experimental value if ESP charges scaled by 0.8 are used. For OPLS, we find, as expected, that unscaled ESP charges give the best results, within 4 kJ/mol of the experimental value. It is interesting that TIP3P water gives results closer to the experimental value than SPC/E. For GROMOS87, ESP charges scaled by 0.6 give results within 4 kJ/mol of the experimental value. The results for GROMOS–690 and GROMOS–POL are similar to those for OPLS and, accordingly, give the best results when unscaled ESP charges are used.

For a certain force field, the LJ energies of the four compounds are similar because the com-

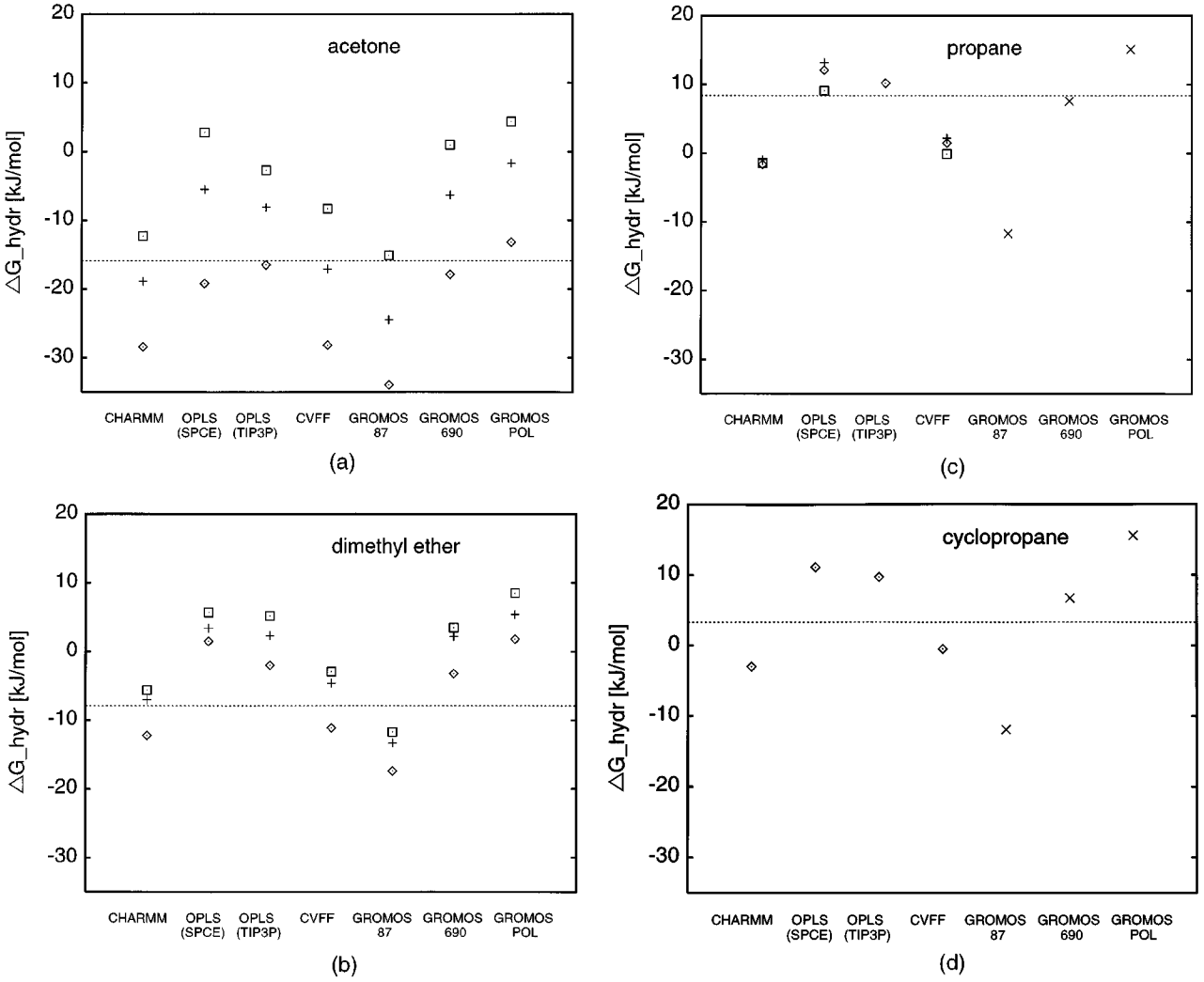

FIGURE 4. ΔG_{hydr} for acetone, dimethylether, propane, and cyclopropane with symbol definitions as in Figure 3. The horizontal dotted lines indicate the experimental values.[52]

pounds have about equal size. Comparison of Figures 3 and 4 shows that the total interaction energies and ΔG_{hydr} values follow similar trends. This observation has been used in the formulation of an extended linear response method.[54]

Figure 5 displays the calculated ΔG_{hydr} values for camphor. The camphor calculations are computationally more intensive due to the larger size of camphor and, consequently, the larger size of the simulated water box. Therefore, the camphor simulations were performed as *initial simulations*. Some of the simulations are not fully converged, e.g., the CHARMM results for 0.6 × ESP and 0.8 × ESP are too close, but their quality is sufficient to show the main trends. Only the simulations with GROMOS–690 were performed as *extended simulations*, and very good convergence was obtained for these (data not shown). The experimental ΔG_{hydr} for camphor of −14.8 kJ/mol is derived from the vapor pressure[55] and the solubility in water.[56] The general trends in simulations for camphor with the different force fields are similar to those observed for the small molecules. The GROMOS–690 results with 0.6 × ESP charges and template charges are slightly too unfavorable (−11.3 kJ/mol and −10.9 kJ/mol, respectively) and the result with 0.8 × ESP charges is slightly too favorable (−17.3 kJ/mol). The magnitude of the GROMOS template charges on the camphor carbonyl group corresponds to ESP charges scaled by 0.68. The ΔG_{hydr} calculated with the GROMOS template charges does not fall between the ΔG_{hydr} calculated with 0.6 × ESP and 0.8 × ESP charges. This is because of the different treatment of the aliphatic moiety in which all atoms

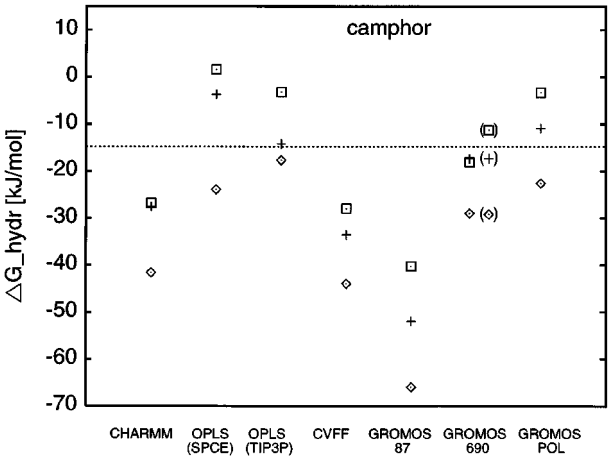

FIGURE 5. ΔG_{hydr} for camphor, with symbol definitions as in Figure 3. All values were computed from *initial simulations* except those shown in parentheses which were computed from *extended simulations* with the GROMOS–690 parameters. The horizontal dotted line indicates the experimental value.

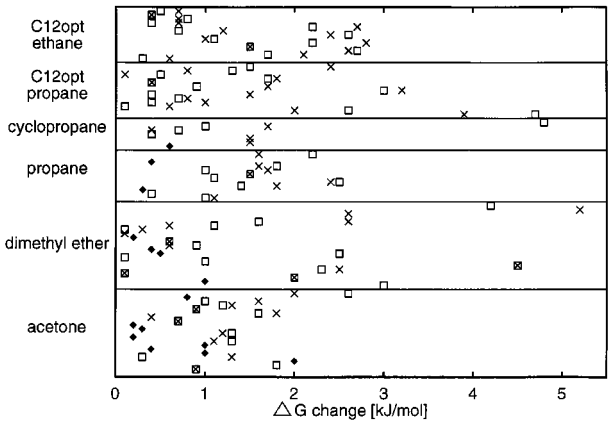

FIGURE 6. The change in ΔG_{hydr} between the *initial simulations* and longer simulations. Simulations are grouped into separate boxes that, from bottom to top, represent simulations for acetone, dimethylether, propane, cyclopropane, C12 optimizations for propane, and C12 optimizations for ethane. Each simulation consisted of 21 windows. Windows of *initial simulations* contained 1000 equilibration and 5000 data acquisition steps (+). Longer windows contained 1000 + (5000 < x < 30,000) steps (♦), 6000 + 15,000 steps (□), or 6000 + (15,000 < x < 30,000) steps (×).

are assigned zero charges in the GROMOS template model. Most of camphor is hydrophobic so the improvement relative to the standard GROMOS87 model is largely due to modification of the carbon–OW LJ parameter.

ACCURACY OF ΔG_{HYDR} CALCULATIONS

The ΔG_{hydr} values for the OPLS force field calculated in this study deviate slightly from published ones[38]: we calculated ΔG_{hydr} for acetone to be more favorable by 1.5 kJ/mol (and closer to the experimental value), and ΔG_{hydr} for dimethylether to be more unfavorable by 3.9 kJ/mol (and further from the experimental value). Orozco et al.[50] calculated ΔG_{hydr} for acetone to be more favorable by 5.5 kJ/mol than Carlson et al.[38] using the same parameters. Our ΔG_{hydr} for propane is more unfavorable by 2.3 kJ/mol (and further from the experimental value) than in a recent calculation where OPLS template charges of −0.18 e and −0.12 e were used for CH_3 and CH_2.[57] These differences can be expected given the statistical errors in the methods used. They may also be partially due to a different cutoff scheme (sharp cutoff vs. quadratic feathering); the use of the TIP3P water model instead of the TIP4P water model; and, in the case of acetone and propane, the use of slightly different atomic charges. We expect that the different ways to circumvent the origin singularity, potential shifting versus bond shrinking to short values, do not affect the calculated results.

The statistical error in ΔG_{hydr} is calculated from the autocorrelation of the data, and describes the statistical reliability of the result. For the 87 ΔG_{hydr} simulations for the small compounds and for the C12-optimizations for propane and ethane, the statistical error decreases from 1.30 ± 0.16 kJ/mol, on average, for simulations of 21 × (1000 + 5000) steps to 0.80 ± 0.13 kJ/mol for 21 × (6000 + 15,000) and 0.71 ± 0.07 kJ/mol for 21 × [6000 + (15,000 < x < 30,000)]. The observed ratio of the statistical errors for 15,000 step windows and 5000 step windows is in the expected range. For an extension of the window length by a factor of 3, one expects a drop of the statistical error by a factor of $\sqrt{3} = 1.73$, and we obtained a ratio of (1.30 ± 0.16)/(0.80 ± 0.13) = 1.6 ± 0.5.

The ΔG_{hydr} change between initial simulations and longer simulations is shown in Figure 6. Apart from five outliers, ΔG_{hydr} is observed to change by less than 3 kJ/mol upon extending the simulation. Most of this change usually takes place in the last three windows. The average ΔG_{hydr} change of 1.4 kJ/mol is less than the sum of the average statisti-

cal errors of the *initial* and the *extended simulations* of ca. 2.0 kJ/mol and approximately equal to the average statistical error in the *initial simulations*.

If the equilibration time in the initial simulations was insufficient, one could expect a drift of $\partial H/\partial\lambda$ during the data acquisition time because the system would not be in equilibrium at the beginning of data acquisition. When we plotted the drift of $\partial H/\partial\lambda$ during the data acquisition stage of the initial simulation against the observed change in ΔG_{hydr} upon extension of the simulation (data not shown), there was no correlation at all. This indicates that the equilibration time of the initial simulations was long enough to remove any correlation between the drift in $\partial H/\partial\lambda$ and the change in ΔG_{hydr} on extending a simulation.

While the statistical error relates to the convergence of a calculated property from a single simulation, the comparison of results from independent simulations indicates how representative the conformational sampling during one simulation is compared to that of a number of simulations. It has been suggested that a number of short FEP simulations are advantageous over one long simulation.[15] We performed five independent calculations for the acetone test system that were started from five different configurations. The average ΔG_{hydr} for the five initial simulations is 16.3 kJ/mol with an average statistical error of 1.35 kJ/mol. The standard deviation of ΔG_{hydr} between the five simulations is 1.0 kJ/mol. The largest deviation is 2.4 kJ/mol. For the *extended simulations*, the average ΔG_{hydr} is 16.5 kJ/mol with an average statistical error of 0.7 kJ/mol. The standard deviation between the results of the five simulations is 0.5 kJ/mol. The largest deviation of the five extended simulations is 1.3 kJ/mol. The average ΔG_{hydr} of the *initial* and *extended* simulations agree surprisingly well. To compare the two approaches (one long simulation versus several short simulations) in terms of efficiency, one has to take into account the 200-ps simulation time that was used for the independent MD equilibration of each of the five starting configurations. The total simulation time for the five *initial* simulations was $5 \times (200\text{ ps} + 252\text{ ps}) = 2.26$ ns, compared to 200 ps + ca. 1 ns = ca. 1.2 ns for one *extended* simulation.

As a second example for studying the behavior of independent simulations, we consider the C12-optimization calculations for ethane and propane, for which we observed a linear relationship between the C12 parameter and the calculated ΔG_{hydr}. Assuming this relationship holds, we performed a linear fit and calculated the deviation of the individual *extended simulation* calculations from this ideal line. This gives another estimate of the deviation of individual calculations. The standard deviation of the simulations with C12 between 650 and 750 [kcal Å^{-12} $\text{mol}^{-1}]^{1/2}$ from the ideal line is 0.9 kJ/mol for ethane and 0.8 kJ/mol for propane. This value is slightly higher than that derived for the acetone system as described in the previous paragraph (0.5 kJ/mol). This might reflect the need for longer simulation times to converge purely hydrophobic interactions. Nevertheless, the small deviation of the results of independent calculations indicates that the relaxation processes in this simple system can be sufficiently and reliably sampled in one long simulation.

To derive an estimate for the accuracy of ΔG_{hydr} calculations from *extended simulations*, one may consider the average statistical error of a single simulation (0.7 kJ/mol), and the difference between independent MCTI calculations (0.5 kJ/mol for the acetone test system and 0.8 kJ/mol for the C12-optimization simulations for ethane and propane). It is unclear whether the statistical error in one trajectory and the deviation between separate trajectories are independent or not. An upper statistical error bound of 1.5 kJ/mol is obtained for *extended simulations* of 0.8–1.0 ns length by simply adding both contributions. A similar procedure can be applied to the shorter *initial* simulations. The average statistical error in one trajectory for the *initial* simulations is 1.3 kJ/mol. The average change in ΔG_{hydr} between an *initial* simulation and an *extended* simulation is 1.4 kJ/mol and the average deviation between separate *extended* simulations is 0.8 kJ/mol. Together, this gives an upper statistical error estimate of $(1.3 + 1.4 + 0.8 =)$ 3.5 kJ/mol for an *initial* simulation of 252 ps in length. This is comparable to the estimated error of 2.7 kJ/mol that was previously derived for a number of MCTI calculations of comparable length for ΔG_{hydr} of SPC/E water.[51]

In addition to statistical errors, there are also systematic contributions to the error, such as those explored here for the C12 parameter. Further improvements in force fields can be expected to reduce these. For example, it has been shown that polarization effects can account for 10–20% of interaction energies,[58] but explicit atomic polarizabilities have only been used in a few cases so far.

Another important factor is the appropriateness of the combining rules for the LJ parameters.[59]

Conclusions

We have shown that absolute ΔG_{hydr} values of small organic molecules can be calculated reliably from MD simulations. The accuracy of ΔG_{hydr} is critically related to the simulation length and the solute size. By analyzing a large number of similar calculations, we have derived estimates for the intrinsic statistical error of a computed ΔG_{hydr} by combining statistical errors from individual simulations, deviations between independent simulations, and the ΔG_{hydr} change upon extension of the simulations. For the set of small compounds studied here, the upper statistical error estimates are 3.5 kJ/mol for 252-ps 21-window simulations and 1.5 kJ/mol for ca. 1-ns simulations with dynamic windowing. The GROMOS87 force field gives ΔG_{hydr} values that are too favorable for hydrophobic compounds. The agreement with experiment can be much improved for such compounds by modification of the LJ interaction parameter between aliphatic carbon and water oxygen atoms. Our simulations for small polar compounds show, however, that charges of larger magnitude than typical in the GROMOS87 force field are likely to be necessary to achieve good results with the modified GROMOS force field.

Acknowledgments

We express our gratitude to T. P. Straatsma for the provision of the ARGOS program and numerous valuable discussions. We thank F. Achard and F. Naroli for implementing a parallel driver for free energy calculations with the ARGOS program. We also thank A. Mark and R. Mancera for helpful discussions, A. Mark and A. Ortiz for critical reading of the manuscript, and A. MacKerell for providing the academic release of the CHARMM 22 parameter set. We thank the HLRZ, Juelich, for access to computing facilities. This work was supported in part by the EU Biotechnology Program (BIO2 CT94-2060). The numbers represented graphi-cally in Figures 3–5 will be accessible via: http://www.embl-heidelberg.de/ExternalInfo/wade/data.html.

References

1. P. E. Smith and B. M. Pettitt, *J. Phys. Chem.*, **98**, 9700 (1994).
2. D. Sitkoff, K. A. Sharp, and B. Honig, *J. Phys. Chem.*, **98**, 1978 (1994).
3. J. Tomasi and M. Persico, *Chem. Rev.*, **94**, 2027 (1994).
4. W. C. Still, A. Tempczyk, R. C. Hawley, and T. Hendrickson, *J. Am. Chem. Soc.*, **112**, 6127 (1990).
5. V. Mohan, M. E. Davis, J. A. McCammon, and B. M. Pettitt, *J. Phys. Chem.*, **96**, 6428 (1992).
6. M. Orozco, W. L. Jorgensen, and F. J. Luque, *J. Comput. Chem.*, **14**, 1498 (1993).
7. S. W. Rick and B. J. Berne, *J. Am. Chem. Soc.*, **116**, 3949 (1994).
8. H. Liu, F. Müller-Plathe, and W. F. van Gunsteren, *J. Chem. Phys.*, **102**, 1722 (1995).
9. U. C. Singh, F. K. Brown, P. A. Bash, and P. A. Kollman, *J. Am. Chem. Soc.*, **109**, 1607 (1987).
10. B. G. Rao and U. C. Singh, *J. Am. Chem. Soc.*, **111**, 3125 (1989).
11. Y. Sun, D. Spellmeyer, D. A. Pearlman, and P. A. Kollman, *J. Am. Chem. Soc.*, **114**, 6798 (1992).
12. W. L. Jorgensen and T. B. Nguyen, *J. Comput. Chem.*, **14**, 195 (1993).
13. R. W. Zwanzig, *J. Chem. Phys.*, **22**, 1420 (1954).
14. W. L. Jorgensen, J. F. Blake, and J. K. Buckner, *Chem. Phys.*, **129**, 193 (1989).
15. C. Chipot, C. Millot, B. Maigret, and P. A. Kollman, *J. Phys. Chem.*, **98**, 11362 (1994).
16. R. C. Wade, M. H. Mazor, J. A. McCammon, and F. A. Quiocho, *J. Am. Chem. Soc.*, **112**, 7057 (1990).
17. D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.*, **94**, 4532 (1990).
18. L. Wang and J. Hermans, *J. Chem. Phys.*, **100**, 9129 (1994).
19. M. Zacharias, T. P. Straatsma, and J. A. McCammon, *J. Chem. Phys.*, **100**, 9025 (1994).
20. T. C. Beutler, A. E. Mark, R. C. van Schaik, P. R. Gerber, and W. F. van Gunsteren, *Chem. Phys. Lett.*, **222**, 529 (1994).
21. T. P. Straatsma and J. A. McCammon, *J. Chem. Phys.*, **91**, 3631 (1991).
22. W. F. van Gunsteren and H. J. C. Berendsen, *Groningen Molecular Simulation (GROMOS) Library Manual*, Biomos, Groningen, The Netherlands (1987).
23. A. R. van Buuren, S. J. Marrink, and H. J. C. Berendsen, *J. Phys. Chem.*, **97**, 9206 (1993).
24. A. E. Mark, S. P. van Helden, P. E. Smith, L. H. M. Janssen, and W. F. van Gunsteren, *J. Am. Chem. Soc.*, **116**, 6293 (1994).
25. P. H. Huenenberger, A. E. Mark, and W. F. van Gunsteren, *J. Mol. Biol.*, **252**, 492 (1995).
26. A. E. Mark (personal communication); X. Daura, P. H. Huenenberger, A. E. Mark, E. Quesol, F. X. Avilés, and W. F. van Gunsteren, *J. Am. Chem. Soc.*, **118**, 6285 (1996).
27. S. Miyamoto and P. A. Kollman, *Proteins*, **16**, 226 (1993).
28. QUANTA version 4.0 parameter handbook, Molecular Simulation Inc., 200 Fifth Avenue, Waltham, MA 02154 (1992).
29. W. L. Jorgensen and J. Tirado-Rives, *J. Am. Chem. Soc.*, **110**, 1657 (1988).

30. P. Dauber-Osguthorpe, V. A. Roberts, D. J. Osguthorpe, J. Wolff, M. Genest, and A. T. Hagler, *Proteins*, **4**, 31 (1988).
31. J. W. Essex, C. A. Reynolds, and W. G. Richards, *J. Am. Chem. Soc.*, **114**, 3634 (1992).
32. V. Helms, E. Deprez, E. Gill, C. Barret, G. Hui Bon Hoa, R. C. Wade, *Biochemistry*, **35**, 1485 (1996).
33. W. D. Cornell, P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. E. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman, *J. Am. Chem. Soc.*, **117**, 5179 (1995).
34. U. C. Singh and P. A. Kollman, *J. Comput. Chem.*, **5**, 129 (1984).
35. B. H. Besler, K. M. Merz, Jr., and P. A. Kollman, *J. Comput. Chem.*, **11**, 431 (1990).
36. C. M. Breneman and K. B. Wiberg, *J. Comput. Chem.*, **11**, 361 (1990).
37. C. I. Bayly, P. Cieplak, W. D. Cornell, and P. A. Kollman, *J. Phys. Chem.* **97**, 10269 (1993).
38. H. A. Carlson, T. B. Nguyen, M. Orozco, and W. L. Jorgensen, *J. Comput. Chem.*, **14**, 1240 (1993).
39. M. Mazor and B. M. Pettitt, *Mol. Sim.*, **6**, 1 (1991).
40. F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. M. J. Smith, M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, and T. Tasumi, *J. Mol. Biol.*, **112**, 535 (1977).
41. Gaussian 92, Revision A, M. J. Frisch, G. W. Trucks, M. Head-Gordon, P. M. W. Gill, M. W. Wong, J. B. Foresman, B. G. Johnson, H. B. Schlegel, M. A. Robb, E. S. Replogle, R. Gomperts, J. L. Andres, K. Raghavachari, J. S. Binkley, C. Gonzalez, R. L. Martin, D. J. Fox, D. J. Defrees, J. Baker, J. J. P. Stewart, and J. A. Pople, Gaussian, Inc., Pittsburgh, PA (1992).
42. H. J. C. Berendsen, J. R. Grigera, and T. P. Straatsma, *J. Phys. Chem.*, **91**, 6269 (1987).
43. K. F. Lau, H. E. Alper, T. S. Thacher, and T. R. Stouch, *J. Phys. Chem.*, **98**, 8785 (1994).
44. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, M. L. Klein, *J. Chem. Phys.*, **79**, 926 (1983).
45. T. P. Straatsma and J. A. McCammon, *J. Comput. Chem.*, **11**, 943 (1990).
46. J. P. Ryckaert, G. Ciccotti, and H. J. C. Berendsen, *J. Comput. Phys.*, **23**, 327 (1977).
47. H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak, *J. Chem. Phys.*, **81**, 3684 (1984).
48. T. P. Straatsma, H. J. C. Berendsen, and A. J. Stam, *Mol. Phys.*, **57**, 89 (1986).
49. D. A. Pearlman and P. A. Kollman, *J. Chem. Phys.*, **90**, 2460 (1989).
50. M. Orozco, W. L. Jorgensen, and F. J. Luque, *J. Comput. Chem.*, **14**, 1498 (1993).
51. V. Helms and R. C. Wade, *Biophys.*, **69**, 810 (1995).
52. S. Cabani, P. Gianni, V. Mollica, and L. Lepori, *J. Solut. Chem.*, **10**, 563 (1981).
53. S. L. Njo, W. F. van Gunsteren, and F. Müller-Plathe, *J. Chem. Phys.*, **102**, 6199 (1995).
54. H. A. Carlson and W. L. Jorgensen, *J. Phys. Chem.*, **99**, 10667 (1995).
55. M. P. Datin, *Ann. Phys.*, **5**, 218 (1916).
56. N. Ishizaka, *Naunyn-Schmiedeberg's Arch. Exp. Pathol. Pharmakol.*, **75**, 194 (1914).
57. G. Kaminski, E. M. Duffy, T. Matsui, and W. L. Jorgensen, *J. Phys. Chem.*, **98**, 13077 (1994).
58. J. Gao, and X. Xia, *Science*, **258**, 631 (1992).
59. M. Waldman and A. T. Hagler, *J. Comput. Chem.*, **14**, 1077 (1993).